

Here I define the Bayesian model I want to use to test the performance of the different priors. The features I want to include are the following: continuous outcome, any number/measurement scale.

Let's start with the **model**

$$\begin{aligned} y_{ij} &= \mathbf{x}_{ij}^T \boldsymbol{\theta} + \mathbf{z}_{ij}^T \mathbf{b}_i + \epsilon_{ij} \\ \mathbf{b}_i &\sim N(\mathbf{0}, \boldsymbol{\Psi}) \\ \epsilon_{ij} &\sim N(0, \sigma^2) \end{aligned} \tag{1}$$

(Vectors are in bold, matrix are capital Greek letters).

I'm going to define the following **priors**:

$$p(\boldsymbol{\theta}) \propto 1 \tag{2}$$

$$p(\sigma^2) \propto \sigma^{-2} \tag{3}$$

For what concerns the random effects variance covariance matrix, different priors are tested. In particular we used:

- inverse-Wishart

$$p(\boldsymbol{\Psi}) \propto IW(\nu_0, S_0) \tag{4}$$

where we choose ν_0 to be 1 and S_0 to be $\text{diag}(2)$, following indications by Gelman *et al* (2014). Thanks to the partitioning property, we know that the diagonal components of $\boldsymbol{\Psi}$ have an inverse-Wishart distribution themselves. Furthermore, for a univariate case, we know that an inverse Wishart distribution simplifies to an inverse Gamma with parameters $p = 1, \alpha = \frac{\nu}{2}, \beta = S_{0kk}$. Finally, we can then apply Jacobian to find the corresponding prior distribution on the standard deviation of any random effect σ_k .

- inverse-Wishart *a la* Huang and Wand

$$\begin{aligned} p(\boldsymbol{\Psi} | \mathbf{a}_1, \mathbf{a}_2) &\propto IW(\nu_0 + p - 1, 2\nu_0 \text{diag}(1/a_1, 1/a_2)), \\ a_k &\propto IG(1/2, 1/A_k^2), \end{aligned} \tag{5}$$

with $\nu_0 = 2$ and $\mathbf{A} = [100, 100]$. Considering a 2×2 random effects variance covariance matrix (random intercepts, and random slopes) that has this prior distribution, the corresponding marginal distributions of the standard deviation random components have an half-Cauchy distribution (belonging to the half- t family, see Gelman (2006) for details). In particular, any square-root diagonal element σ_k of a matrix following such a distribution is Half- $t(\nu_0, A_k)$. When, $\nu = 1$, and A_k is equal to the prior guess we have an half-Cauchy prior distribution for a σ_k parameter. Furthermore, the marginal priors for the correlations have a beta distribution with parameters $\alpha = -1, \beta = -1$ (Huang and Wand, 2013).

- Matrix-F variate

$$\begin{aligned} p(\boldsymbol{\Psi}) &\propto F(\boldsymbol{\Psi}; \nu_0, \delta, \mathbf{B}) \\ &\propto \int IW(\boldsymbol{\Psi}; \delta + k - 1, \Sigma) \times W(\Sigma; \nu_0, \mathbf{B}) d\Sigma \end{aligned} \tag{6}$$

where $\nu_0 = 2$, $\delta = 1$, and \mathbf{B} is a prior guess. Three different choices were made for \mathbf{B} in this paper: $\text{diag}(10^3)$, proper neighbor of $(\sigma^2)^{-\frac{1}{2}}$; \mathbf{B}_{ed} , an educated guess based on data exploration, \mathbf{R}^* and an empirical bayes choice following Kass and Natarajan (2006).

Considering a 2×2 random effects variance covariance matrix (random intercepts, and random slopes) that is matrix-F distributed, $F(\nu_0, \delta, \mathbf{B})$, the marginal distribution on the standard deviations of the random effects are univariate $F(\nu_0, \delta, b_{11})$ and $F(\nu_0, \delta, b_{22})$, with $\nu_0 > 1, \delta > 0, b_{jj} > 0$.

The derivation of the conditional posterior follows.

Full conditional for θ (fixed effects)

Let's start with

$$p(\theta|\mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{b}_i, \Psi, \sigma^2) = p(\mathbf{y}|\theta, \mathbf{X}, \mathbf{Z}, \mathbf{b}_i, \Psi, \sigma^2)p(\theta)$$

where

$$\begin{aligned} p(\mathbf{y}|\theta, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2) &= \prod_{i=1}^n \prod_{j=1}^J p(y_{ij}|\theta^T \mathbf{x}_{ij} + \mathbf{b}_i^T \mathbf{z}_{ij}, \Psi, \sigma^2) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} SSR\right) \end{aligned}$$

and

$$SSR = \sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \theta^T \mathbf{x}_{ij} - \mathbf{b}_i^T \mathbf{z}_{ij})^2$$

where can rewrite y_{ij} as \tilde{y}_{ij} , with $\tilde{y}_{ij} = y_{ij} - \mathbf{b}_i^T \mathbf{z}_{ij}$ which makes SSR:

$$\begin{aligned} SSR &= \sum_{i=1}^n \sum_{j=1}^J (\tilde{y}_{ij} - \theta^T \mathbf{x}_{ij})^2 \\ &= (\tilde{\mathbf{y}} - \mathbf{X}\theta)^T (\tilde{\mathbf{y}} - \mathbf{X}\theta) \\ &= \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - 2\theta^T \mathbf{X}\tilde{\mathbf{y}} + \theta^T \mathbf{X}^T \mathbf{X}\theta \end{aligned}$$

Hence,

$$p(\mathbf{y}|\theta, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2} [-2\theta^T \mathbf{X}\tilde{\mathbf{y}} + \theta^T \mathbf{X}^T \mathbf{X}\theta]\right)$$

Combining this with the prior we obtain:

$$\begin{aligned} p(\theta|\mathbf{y}, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2) &\propto \exp\left(-\frac{1}{2}\theta^T \mathbf{X}^T \mathbf{X}\theta + \theta^T \mathbf{X}\tilde{\mathbf{y}}\right) \\ \theta|. &\sim N\left(\frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}\tilde{\mathbf{y}}}{\sigma^2}, \frac{(\mathbf{X}^T \mathbf{X})^{-1}}{\sigma^2}\right) \end{aligned} \tag{7}$$

Full conditional for \mathbf{b}_i (random effects)

To derive this one we can start from:

$$p(\mathbf{b}_i|\mathbf{y}, \mathbf{X}, \mathbf{Z}, \theta, \Psi, \sigma^2) = p(\mathbf{y}_i|\theta, \mathbf{b}_i, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2)p(\mathbf{b}_i)$$

We know that

$$p(\mathbf{y}_i|.) = \prod_{j=1}^J p(y_{ij}|\theta^T \mathbf{x}_{ij} + \mathbf{b}_i^T \mathbf{z}_{ij}, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2} SSR_i\right)$$

with

$$SSR = \sum_{j=1}^J (y_{ij} - \theta^T \mathbf{x}_{ij} - \mathbf{b}_i^T \mathbf{z}_{ij})^2$$

and we can rewrite y_{ij} as $\tilde{y}_{ij} = y_{ij} - \theta^T \mathbf{x}_{ij}$, which would make SSR be

$$\begin{aligned} SSR &= \sum_{j=1}^J (\tilde{y}_j - \theta^T \mathbf{x}_j)^2 \\ &= (\tilde{\mathbf{y}} - \mathbf{b}_i^T \mathbf{Z}_i)^T (\tilde{\mathbf{y}} - \mathbf{b}_i^T \mathbf{Z}_i) \\ &= \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - 2\mathbf{b}_i^T \mathbf{Z}_j \tilde{\mathbf{y}} + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{b}_i \end{aligned}$$

Hence,

$$p(\mathbf{y}_i|\cdot) \propto \exp(-\frac{1}{2\sigma^2}[-2\mathbf{b}_i^T \mathbf{Z}_j \tilde{\mathbf{y}} + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{b}_i])$$

We also know that in this case, the "prior" is

$$p(\mathbf{b}_i) \propto N(\mathbf{0}, \Psi) \propto \exp(-\frac{1}{2}[-2\mathbf{b}_i^T \Psi^{-1} \mathbf{0} + \mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i])$$

In conclusion, combining the sampling model and the prior, we get:

$$\begin{aligned} p(\mathbf{b}_i|\cdot) &\propto \exp(-\frac{1}{2\sigma^2}[-2\mathbf{b}_i^T \mathbf{Z}_j \tilde{\mathbf{y}} + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{b}_i] - \frac{1}{2\sigma^2}[-2\mathbf{b}_i^T \mathbf{Z}_j \tilde{\mathbf{y}} + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{b}_i]) \\ \mathbf{b}_i| \cdot &\propto N\left(\left(\Psi^{-1} + \frac{\mathbf{Z}_i^T \mathbf{Z}_i}{\sigma^2}\right)^{-1} \left(\Psi^{-1} \mathbf{0} + \frac{\mathbf{Z}_i^T \tilde{\mathbf{y}}_i}{\sigma^2}\right), \left(\Psi^{-1} + \frac{\mathbf{Z}_i^T \mathbf{Z}_i}{\sigma^2}\right)^{-1}\right) \end{aligned} \quad (8)$$

Full conditional for σ^2 (error variance)

The full conditional posterior can be expressed as:

$$p(\sigma^2|\mathbf{y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \mathbf{b}_i, \Psi) = p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{b}_i, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2)p(\sigma^2)$$

The sampling model is the same we saw for the full conditional distribution of $\boldsymbol{\theta}$:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2) &= \prod_{i=1}^n \prod_{j=1}^J p(y_{ij}|\boldsymbol{\theta}^T \mathbf{x}_{ij} + \mathbf{b}_i^T \mathbf{z}_{ij}, \Psi, \sigma^2) \\ &= \prod_{i=1}^n \prod_{j=1}^J (2\pi\sigma^{-2})^{-\frac{1}{2}} \exp(-\frac{(y_{ij} - \boldsymbol{\theta}^T \mathbf{x}_{ij} - \mathbf{b}_i^T \mathbf{z}_{ij})^2}{2\sigma^2}) \end{aligned}$$

However, we are now interested in σ^2 , hence

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}, \mathbf{Z}, \Psi, \sigma^2) &\propto (\sigma^2)^{-\frac{N}{2}} \exp(-\frac{\sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \boldsymbol{\theta}^T \mathbf{x}_{ij} - \mathbf{b}_i^T \mathbf{z}_{ij})^2}{2\sigma^2}) \\ &\propto (\sigma^2)^{-\frac{N}{2}} \exp(-\frac{1}{2\sigma^2} SSR) \end{aligned}$$

where $N = \sum_i^n n_j$ is the entire sample size (all observations within all clusters). The prior for σ is given above, and therefore we can write the full conditional posterior as:

$$\begin{aligned} p(\sigma^2|\mathbf{y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \mathbf{b}_i, \Psi) &\propto (\sigma^2)^{-\frac{N}{2}-1} \exp(-\frac{1}{2\sigma^2} SSR) \\ \sigma^2| \cdot &\sim IG(\frac{N}{2}, \frac{SSR}{2}) \end{aligned} \quad (9)$$

Full conditional for Ψ (random effects variance covariance matrix)

Here, we need to write down the posteriors for the different priors we specified. First, let us define the sampling model for the random effects.

$$\begin{aligned} \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} &= \mathbf{b}_i \sim N(\mathbf{0}, \Psi) \\ p(\mathbf{b}_1, \mathbf{b}_2|\Psi) &\propto |\Psi|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \text{tr}(\mathbf{S}_b \Psi^{-1})\right) \end{aligned} \quad (10)$$

where \mathbf{S}_b is $\sum_i \mathbf{b}_i \mathbf{b}_i^T$

- given the inverse-Wishart prior

$$p(\Psi) \propto IW(\nu_0, \mathbf{S}_0) \\ \propto |\Psi|^{-\frac{(\nu_0+k+1)}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{S}_0 \Psi^{-1})\right)$$

the full conditional posterior of Ψ is

$$p(\Psi|\cdot) \propto |\Psi|^{-\frac{(\nu_0+k+1+n)}{2}} \exp\left(-\frac{1}{2}\text{tr}([\mathbf{S}_0 + \mathbf{S}_b] \Psi^{-1})\right) \\ \propto IW(\nu_0 + n, \mathbf{S}_0 + \mathbf{S}_b) \quad (11)$$

where $\nu_0 = 2$

- inverse-Wishart *a la* Huang and Wand

$$p(\Psi|a_1, a_2) \propto IW(\nu_0 + k - 1, 2\nu_0 \text{diag}(1/a_1, 1/a_2)), \\ a_k \propto IG(\eta, 1/A_k^2) \\ p(\Psi) \propto |\Psi|^{-\frac{(\nu_0+k-1+1)}{2}} \exp\left(-\frac{1}{2}\text{tr}(2\nu_0 \text{diag}(1/a_1, 1/a_2) \Psi^{-1})\right) \\ \times \left(\frac{1}{a_1}\right)^{\eta+1} \exp\left(-\frac{1}{A_1^2 a_1}\right) \times \left(\frac{1}{a_2}\right)^{\eta+1} \exp\left(-\frac{1}{A_2^2 a_2}\right)$$

the full conditional posterior of Ψ is

$$p(\Psi|\cdot) \propto |\Psi|^{-\frac{(\nu_0+k-1+n+1)}{2}} \exp\left(-\frac{1}{2}\text{tr}([\mathbf{S}_b + 2\nu_0 \text{diag}(1/a_1, 1/a_2)] \Psi^{-1})\right) \\ \propto IW(\nu_0 + k - 1 + n, \mathbf{S}_b + 2\nu_0 \text{diag}(1/a_1, 1/a_2)) \\ p(a_k|\cdot) \propto IG\left(\eta(\nu_0 + k), \nu_0 \left(\Psi_{kk}^{-1} + \frac{1}{A_k^2}\right)\right) \quad (12)$$

where $\eta = \frac{1}{2}$, $\nu_0 = 2$, $k = 2$, and n is the number of clusters (individuals). (For the conditional posterior of a_k refer to Huang and Wand (2013), section 4.2).

- Matrix-F variate

$$p(\Psi) \propto F(\Psi; \nu_0, \delta, \mathbf{B}) \\ \propto \int IW(\Psi; \delta + k - 1, \mathbf{\Omega}) \times W(\mathbf{\Omega}; \nu_0, \mathbf{B}) d\mathbf{\Omega} \\ \propto \int |\Psi|^{-\frac{(\delta+2k)}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{\Omega} \Psi^{-1})\right) \times |\mathbf{\Omega}|^{-\frac{(\nu-k-1)}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{\Omega} \mathbf{B}^{-1})\right) d\mathbf{\Omega}$$

the full conditional posterior of Ψ is

$$p(\Psi|\cdot) \propto \int |\Psi|^{-\frac{(\delta+2k+n)}{2}} \exp\left(-\frac{1}{2}\text{tr}([\mathbf{\Omega} + \mathbf{S}_b] \Psi^{-1})\right) \times |\mathbf{\Omega}|^{-\frac{(\nu-k-1)}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{\Omega} \mathbf{B}^{-1})\right) d\mathbf{\Omega} \\ \propto \int IW(\delta + k + n - 1, \mathbf{\Omega} + \mathbf{S}_b) \times W(\nu, \mathbf{B}) d\mathbf{\Omega} \quad (13)$$

with $\nu = k = 2$, $\delta = 1$ and \mathbf{B} prior guess

Notation Conventions

- n number of clusters; i specific cluster
- J number of observations within cluster; j specific observation
- N total number of observations